# PSCR 2020:

## THE DIGITAL EXPERIENCE

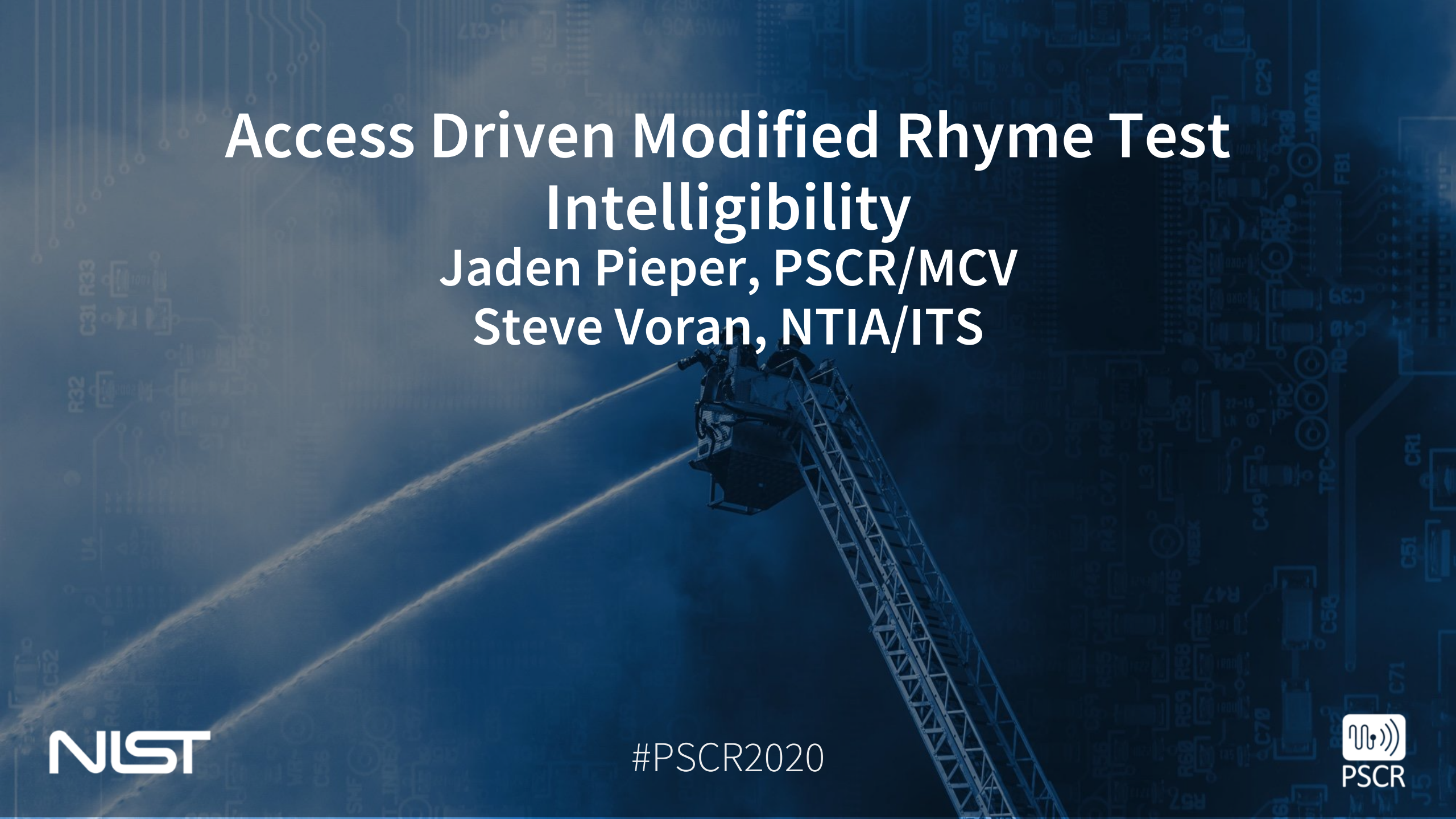
NIST

#PSCR2020

PSCR

# Access Driven Modified Rhyme Test Intelligibility

Jaden Pieper, PSCR/MCV
Steve Voran, NTIA/ITS



NIST

#PSCR2020

PSCR

# DISCLAIMER

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately.

Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

*** Please note, unless mentioned in reference to a NIST Publication, all information and data presented is preliminary/in-progress and subject to change**

**Exempt Human Subjects Research: The National Institute of Standards and Technology Research Protections Office reviewed the protocol for this project and determined it meets the criteria for "exempt human subjects research" as defined in 15 CFR 27, the Common Rule for the Protection of Human Subjects.**

# Overview

- Quality of Experience Based Measurements
- Access Time Measurement System
- Intelligibility
- New Audio Files
- New MRT Application
- Path Forward

# Quality of Experience Based Measurements

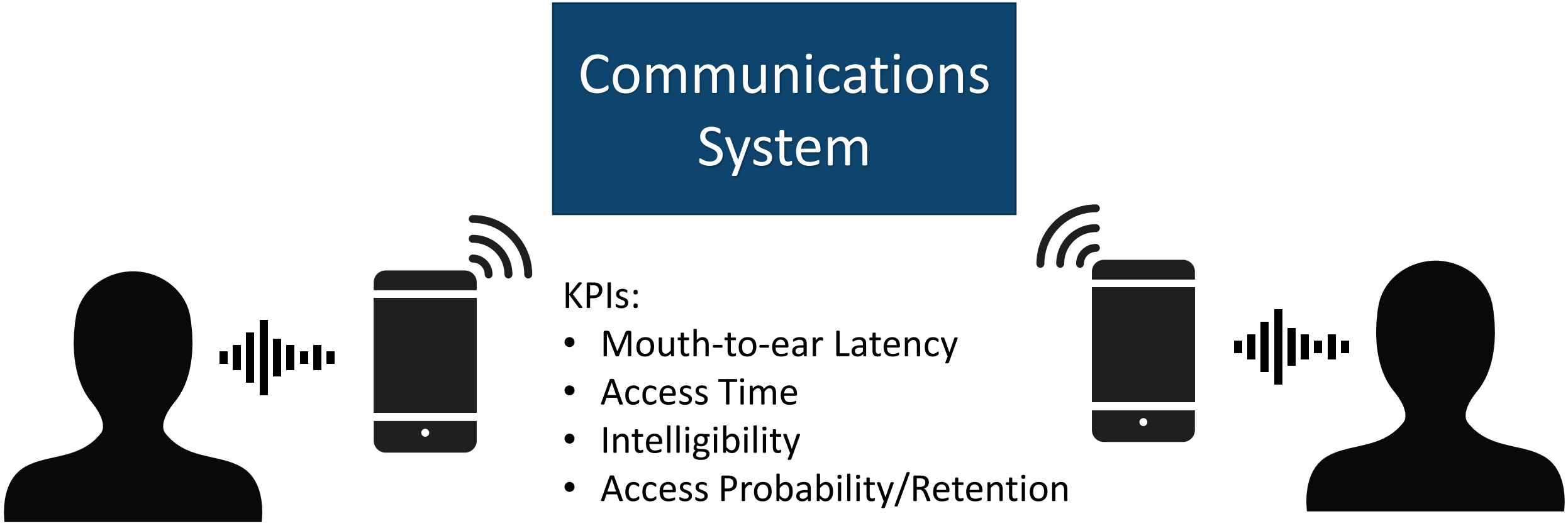
- **QoE KPIs for Mission Critical Voice (MCV)**
  - Mouth-to-Ear Latency
  - Access Time
  - Voice Quality/Intelligibility
  - Access/Retention Probability
- **MCV QoE Measurement Methods Presentation**
- **Mouth-to-Ear Latency Measurement System**
- **Access Time Measurement System**

# The User Experience: PTT Communications

- **Press PTT and speak into a device**
- **Listening to speech output from a device**
- **It's all about speech**

- **Goal - Create measurement systems that are:**
  - Based upon the user experience -- speech
  - Comparable and fair across technologies
- **This is not:**
  - Diagnosing internal components of specific systems

# Technology Agnostic Measurements



# User Driven Access Definition

- **End-to-end Access Time**
  - *The total amount of time from when a transmitting user first presses PTT until a receiving user hears intelligible audio.*
- **Two Components:**
  - Mouth-to-ear Latency
    - *The time between speech being input into one device and its output through another*
  - Access Delay
    - *The minimum length of time a user must wait between pressing a PTT button and starting to speak to ensure that the start of the message is not lost*

# Formalizing Access Delay

- **Access Delay**
  - All about if a message is lost or not
  - Intelligibility is the key to the measurement
- **Formal Definition:**
  - *The minimum length of time a user must wait between pressing a PTT button and starting to speak to ensure that the first word of the message has an average intelligibility that is no lower than* $\alpha \cdot I_0$
  - $0 < \alpha < 1$, defines acceptable intelligibility level
  - $I_0$ is the baseline intelligibility of that word through the communications system
    - No system is perfectly intelligible
    - Some level of degradation almost always present

# Intelligibility

- **Modified Rhyme Test (MRT)**
  - Used to test intelligibility of SCBA masks[1]
  - Batches of six words
    - *went, sent, bent, dent, tent, rent*
    - Words: consonant-vowel-consonant
    - Each batch: Either leading or trailing consonant varies
  - MRT Trial
    - Carrier phrase + word
    - e.g. “Please select the word *went*”
    - Success (identified) or Failure (mis-identified)
  - Over lots of trials scores are generated
    - Score is value between 0 and 1
    - Corrected for guessing
  - High time cost

1: NFPA 1981 Standard on open-circuit self-contained breathing apparatus (SCBA) for emergency services (2007)

# Intelligibility

## MRT: Standard Use Case Examples

**Batch:** ***fun, sun, bun, gun, run, nun***
**Distortion: Background noise + system**

- **Extreme** 
- **Moderate** 
- **Mild** 
- **None** 

# Intelligibility

- **Articulation Band Correlated Modified Rhyme Test (ABC-MRT)**
  - Objective algorithm to provide estimates of true MRT scores
  - Most recent version is ABC-MRT16
  - Relies on temporal correlations within articulation index bands
    - Break speech into a “musical score”
    - Representation of speech in time and frequency
  - Costs: Cheaper and faster
    - Relatively low time cost
      - Can get estimated MRT scores “on demand”
    - Much lower infrastructure cost

# Measurement Concept

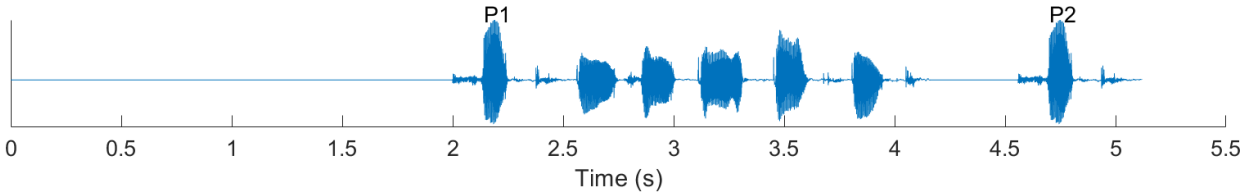
Overview

- **Access Delay definition:**
  - *The minimum length of time a user must wait between pressing a PTT button and starting to speak to ensure that the first word of the message has an average intelligibility that is no lower than* $\alpha \cdot I_0$
- **Repeatedly send pre-defined audio clips through communications system**
- **Vary where in the clip PTT is triggered**
- **Measure relationship between PTT time and intelligibility of the first word in the clip**
  - No more carrier phrase
  - Starting to move outside intended scope of ABC-MRT16

# Measurement Concept

Audio Clips

- **Select single word from ABC-MRT16 database[1]**
  - Use only words from batches where leading consonant varies
    - E.g. *went, sent, bent, dent, tent, rent*
    - Places majority of intelligibility emphasis on beginning of word
- **Structure:**
  - $T$ seconds of silence
  - Play word, $P_1$
  - $T$ seconds of speech
  - Play word again, $P_2$

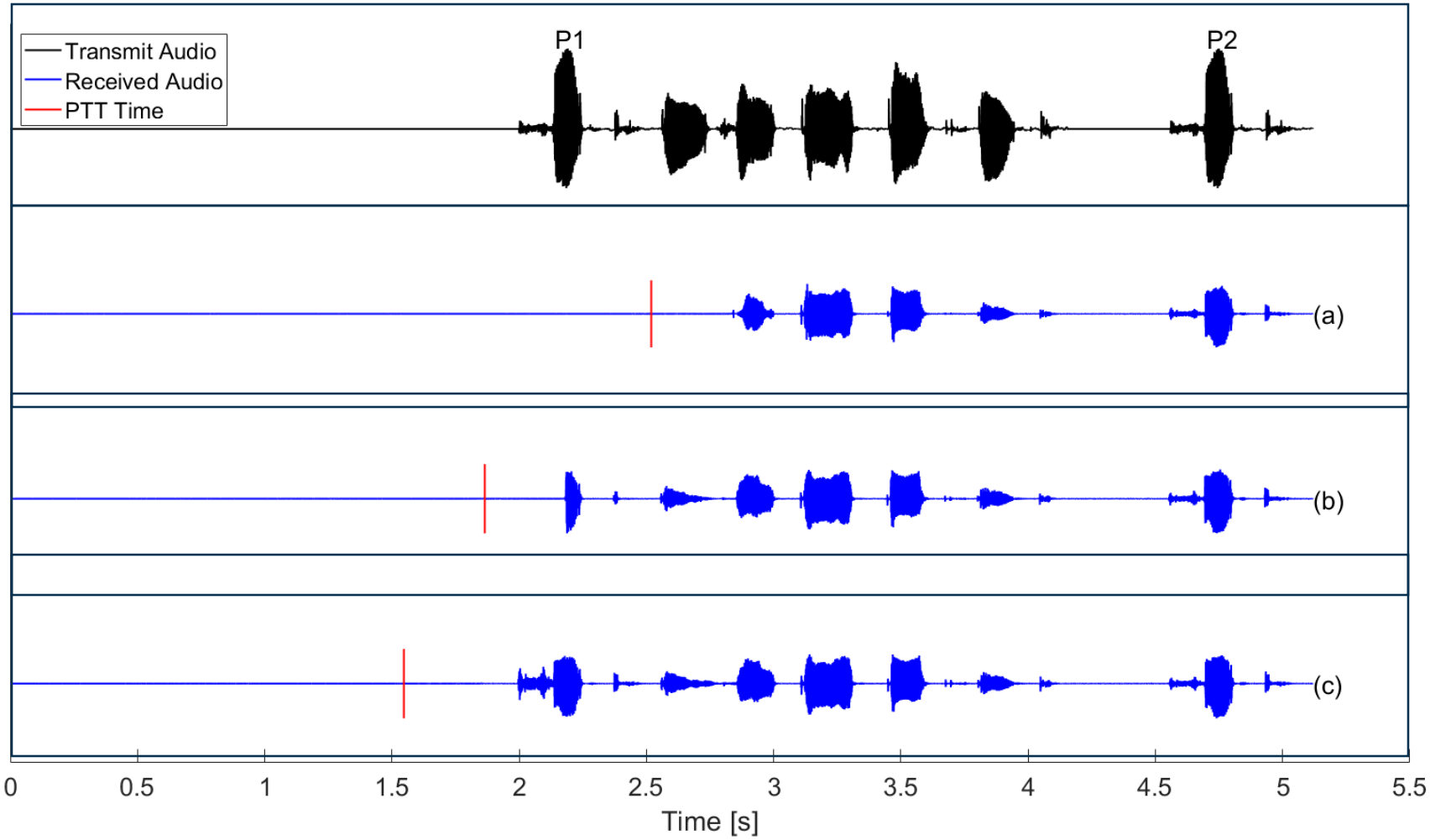



- **$T$ chosen so that system access time is less than $T$ seconds**
- **Intelligibility of $P_2$ describes the asymptotic intelligibility, $I_0$**
- **Intelligibility of $P_1$ relates PTT time with intelligibility**

1: Voran SD (2017) A multiple bandwidth objective speech intelligibility estimator based on articulation index band correlations and attention. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 5100–5104.doi: 10.1109/ICASSP.2017.7953128

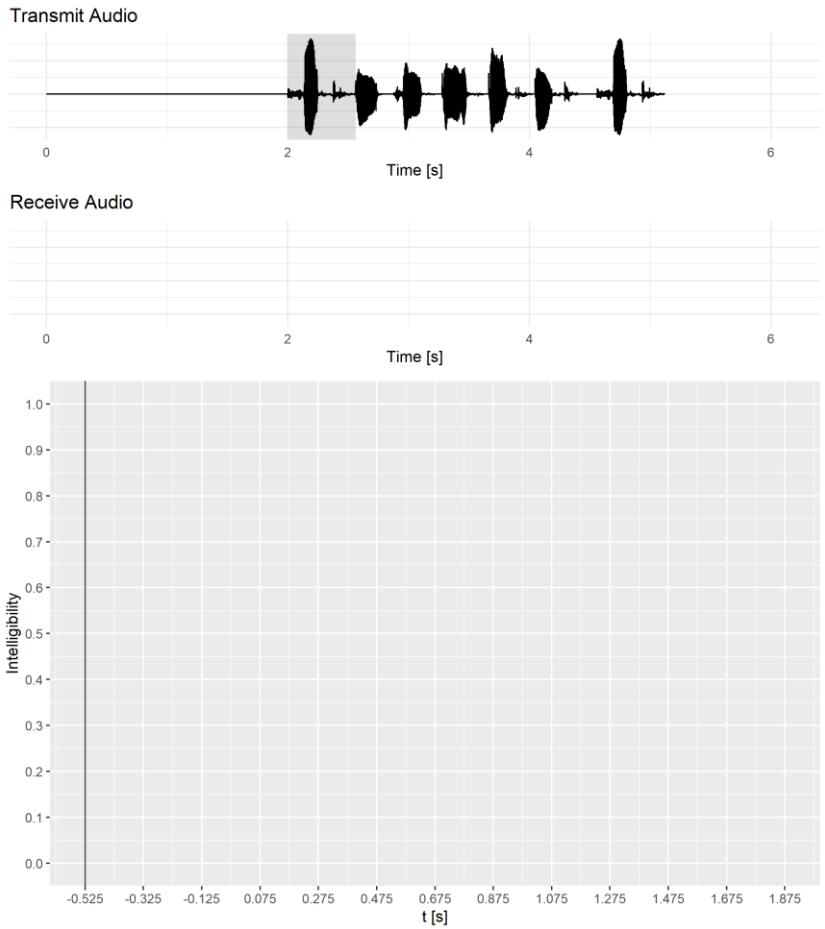
# Measurement Concept

Intelligibility Examples: *hook*



# Access Time

***Play Access Animation Here***



# Measurement Limitations

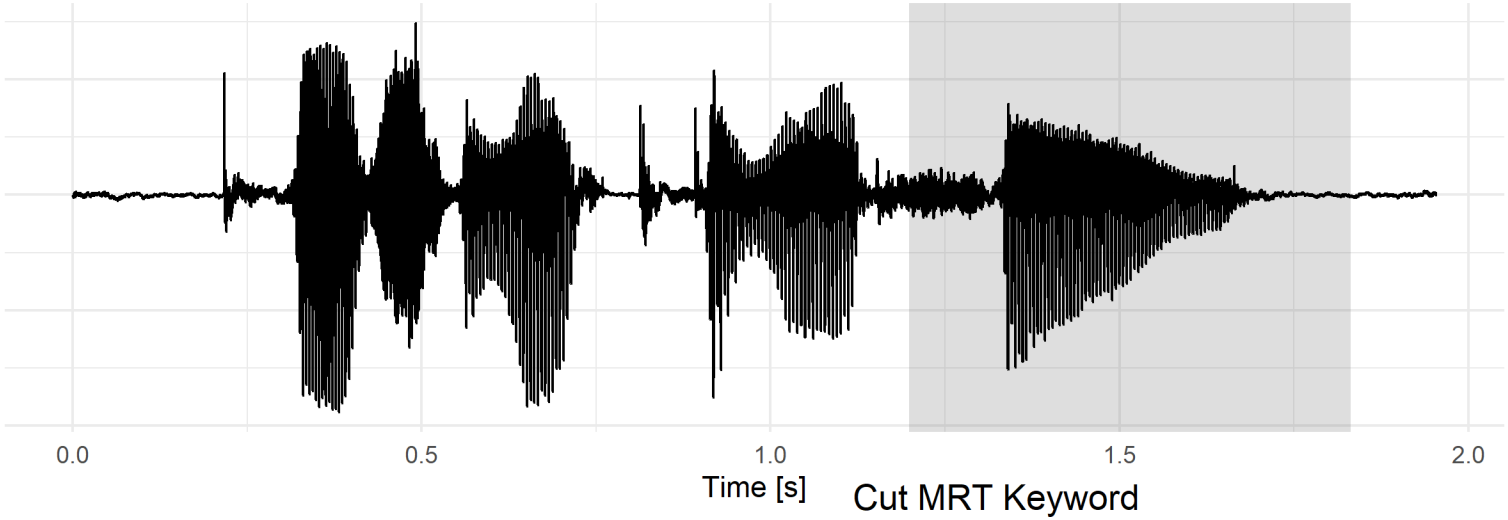
- **MRT keywords not designed for this focus**
  - MRT keywords extracted from full MRT phrases
  - “Please select the word west”
  - Coarticulation
- **ABC-MRT not designed for this impairment**
  - ABC-MRT designed around normal MRT impairments
  - Not designed with partially muted words in mind
  - Data does not exist
  - Can not guarantee response mimics human’s
- **Need to run new MRT experiments**
  - Focus on partially muted words

# Beginning of MRT Keywords

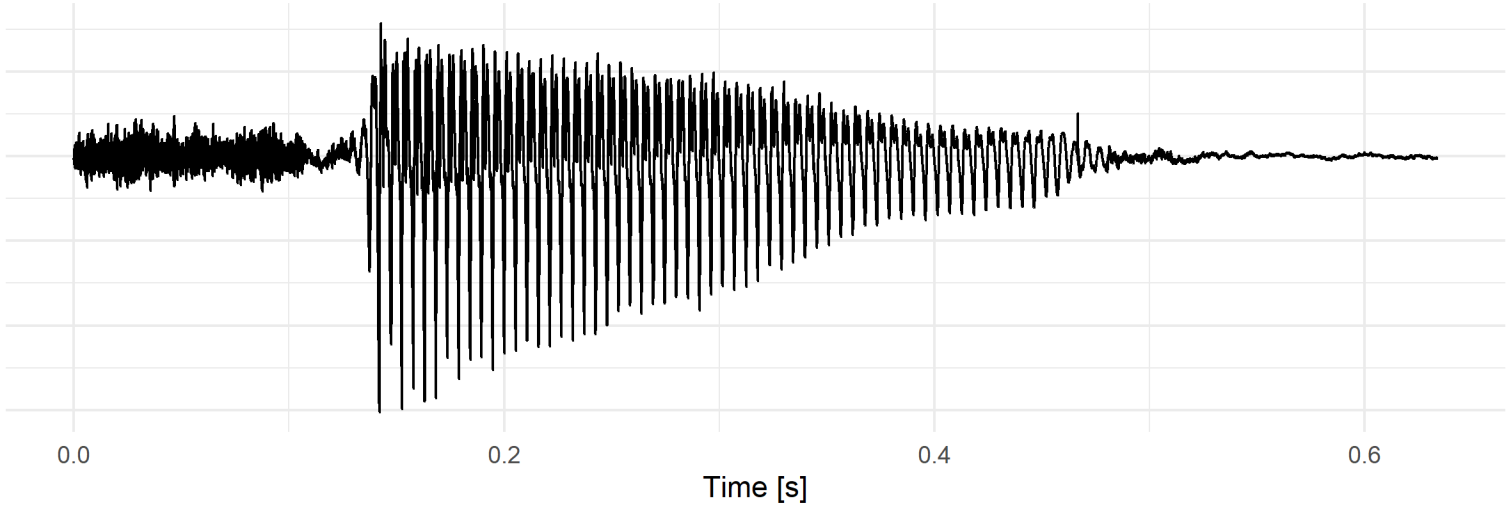
- **Beginning of the word**
  - Key to access delay measurements
  - Built into the definition
    - *The minimum length of time a user must wait between pressing a PTT button and starting to speak to ensure that the first word of the message has an average intelligibility that is no lower than* $\alpha \cdot I_0$
- **How do you determine the start of the keyword?**
  - Extract from recorded MRT speech
  - “Please select the word west”

# Extracting MRT Keywords

Full MRT Phrase

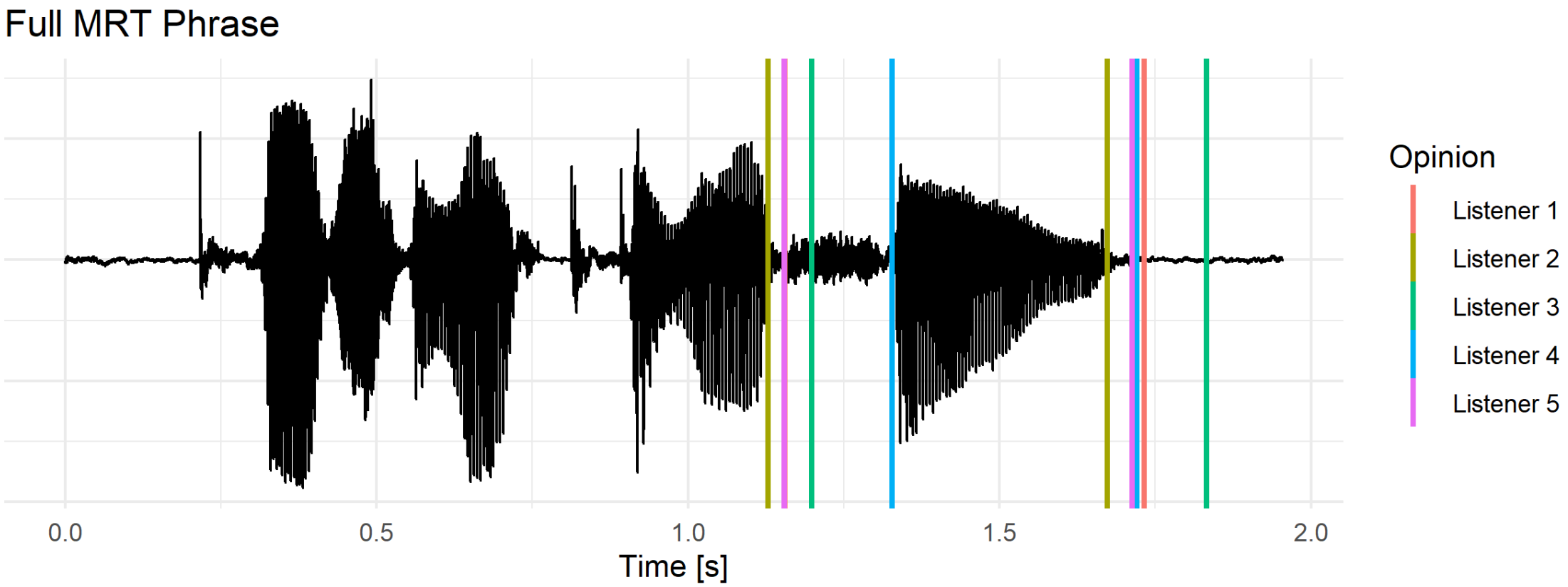



Cut MRT Keyword




# Extracting MRT Keywords

- **Get multiple opinions**



# New Audio

- **Extracting coarticulated speech is hard**
- **Design speech for this problem from the ground up**
  - Isolate keywords
  - Knowledge of when a talker is speaking

# New Audio

# New Audio

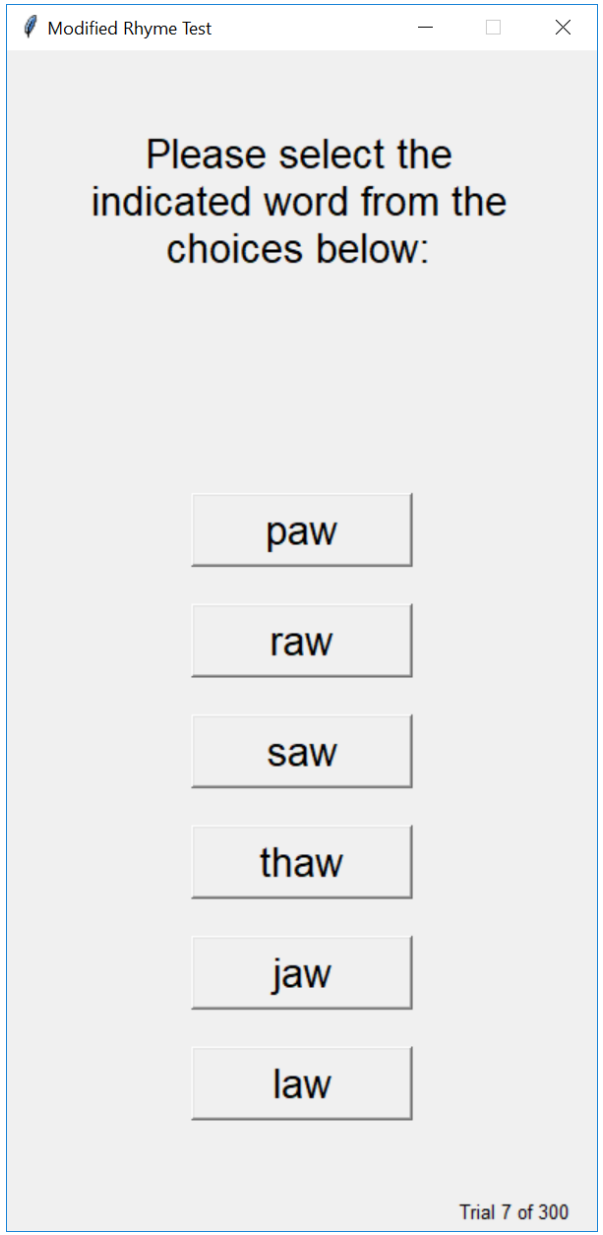
- **Four talkers**
- **300 MRT keywords**
  - 1 second of “silence” between keywords
- **Carrier phrase 3 times**
  - Beginning, middle, end
- **List of 10 Harvard Phrases**
  - Phonetically balanced
  - Good audio for mouth-to-ear latency measurements
  - “Hoist the load to your left shoulder.”

# New Audio: Extracting MRT Keywords

- **Talker PTT**
  - Identify when talkers think word is spoken
  - Quick reference for extracting MRT keywords
  - Large gaps between PTT and speech
  - Limited scope for conclusions
- **Manual Corrections**
  - Talkers made mistakes
  - Verify each word

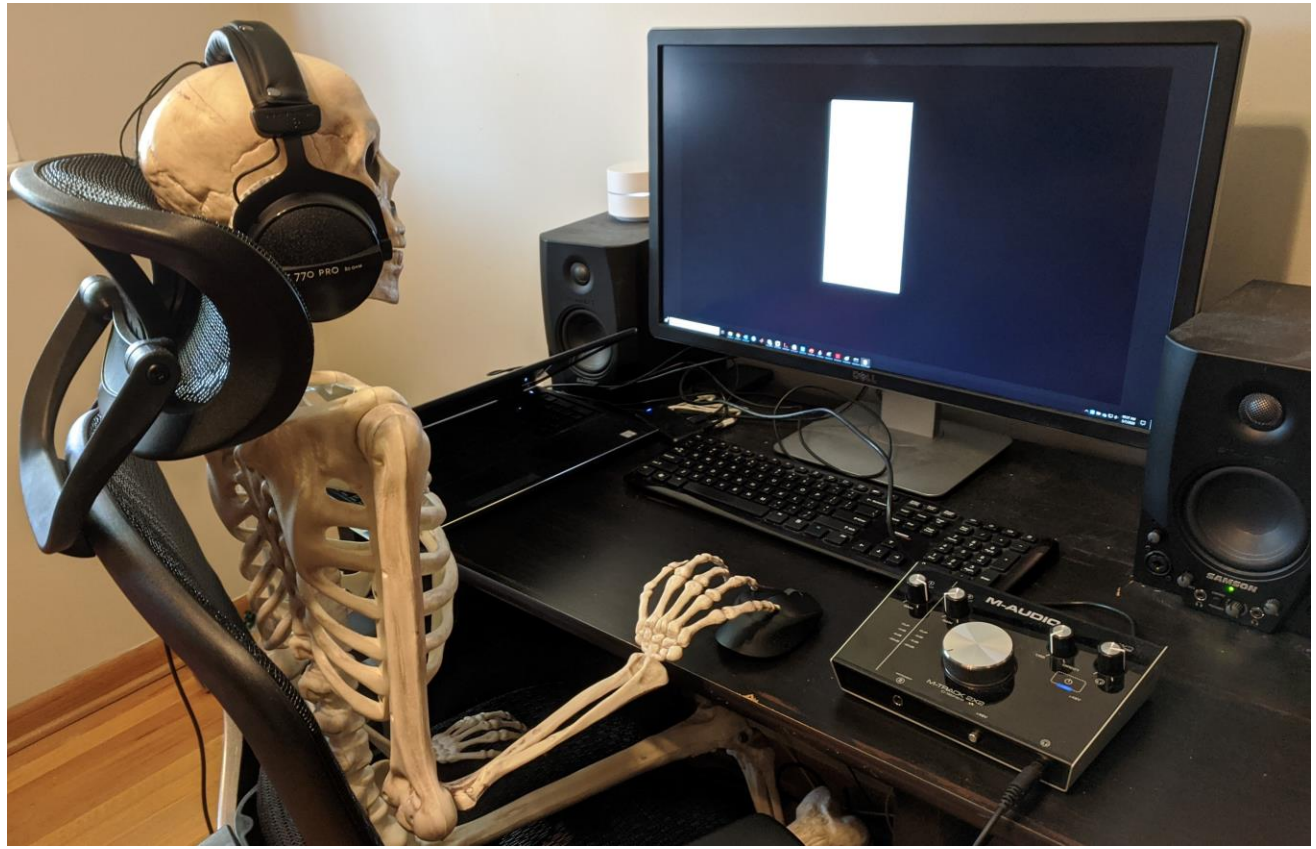
# MRT Application

- **ITS MATLAB GUI**
  - Built in 2007
  - Requires license
- **Open source MRT application**
  - Could not find any
- **Design new MRT application**
  - Free and easy to use
  - Python



# Verify Baseline Intelligibility

- **Test Design**
  - Five Listeners
  - 4 sessions each
  - 300 trials per session
- **Components**
  - Isolated room
  - Sound isolation - closed studio headphones
  - Audio interface



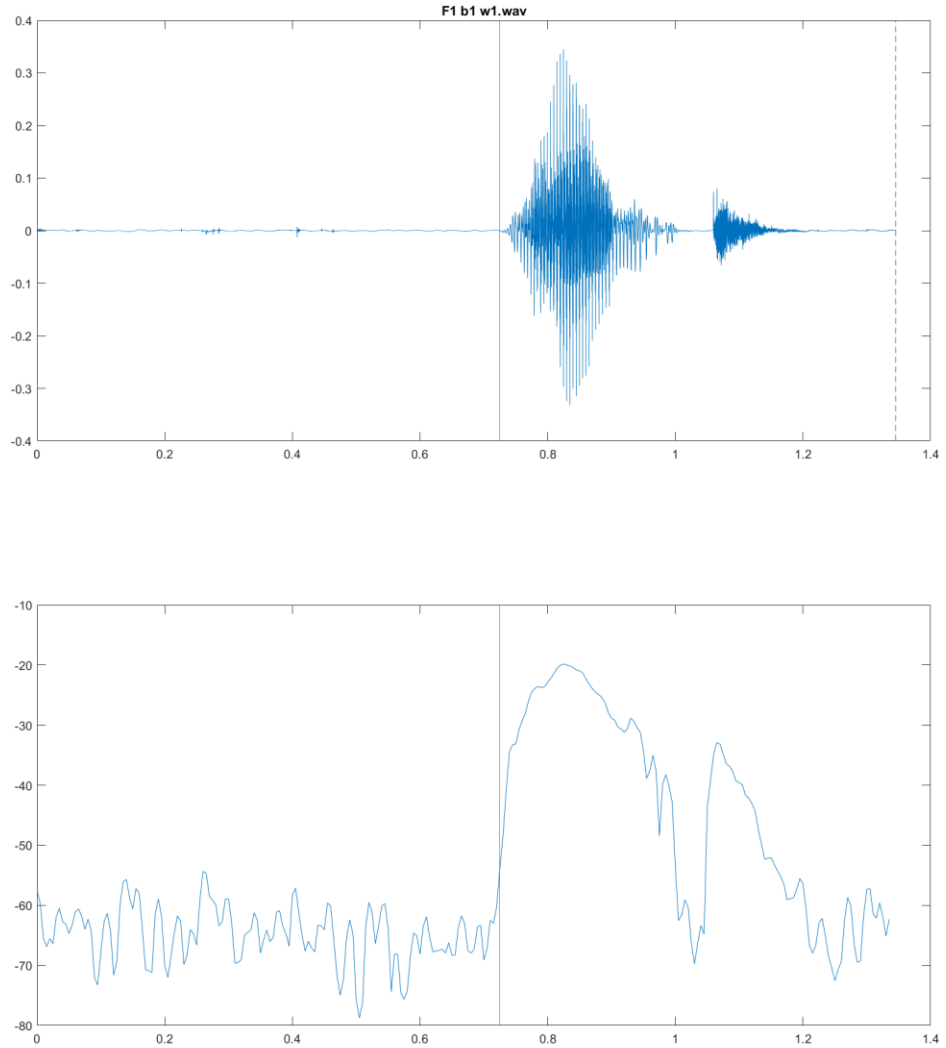
# Verify Baseline Intelligibility

- **Baseline intelligibility: 99.1%***
  - *Audio issues
- **Expect results to improve**

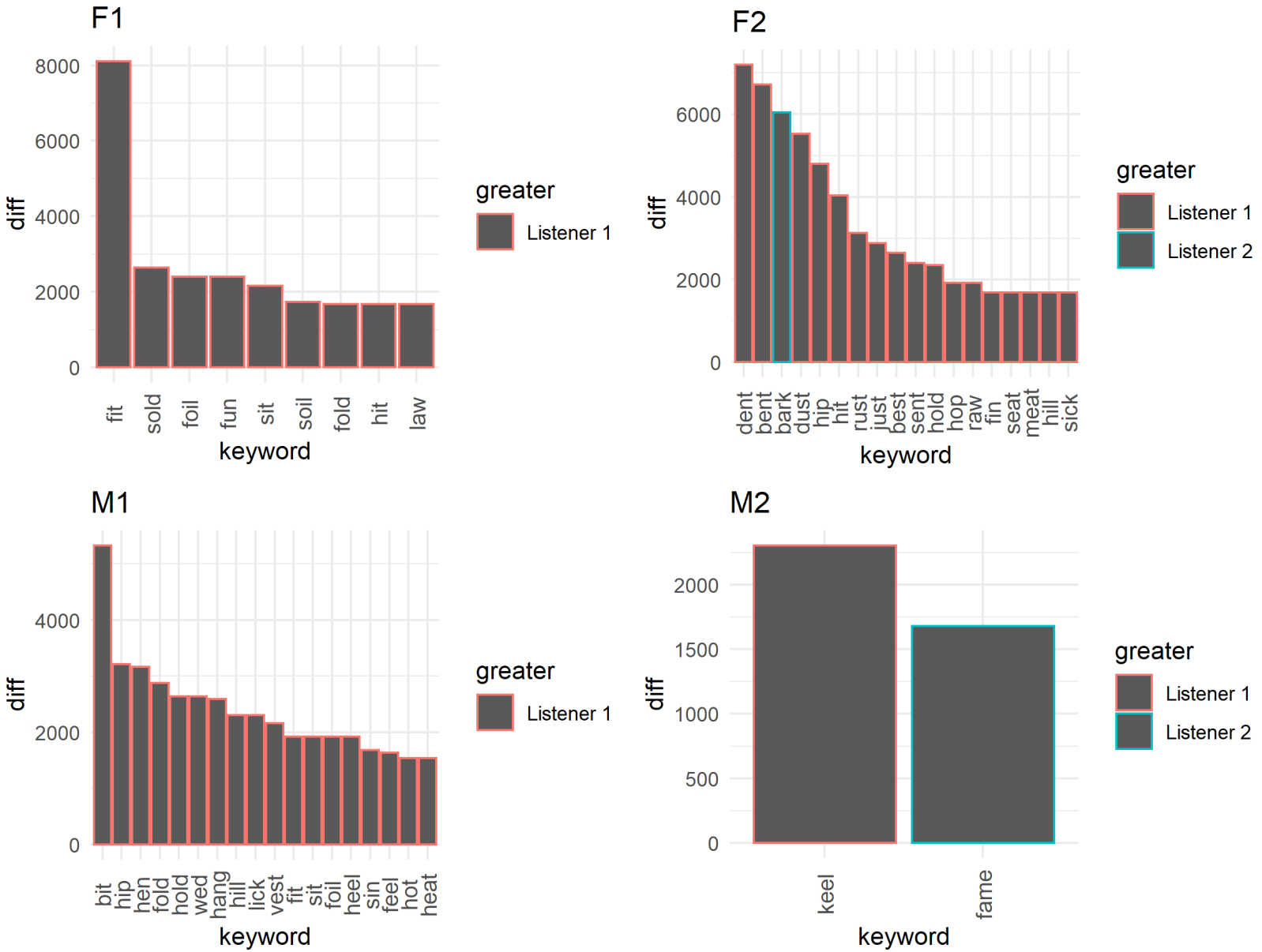
| Talker | Score |
|---|---|
| F1 | 99.0% |
| F2 | 99.1% |
| M1 | 98.9% |
| M2 | 99.3% |

# Identify Start of Audio

- **Accuracy decreases cost of human MRT trials**
- **Simple MATLAB application**
  - Process much faster
- **Essentially quantizes selections**
  - Should cause more agreement
- **Still disagreement**



# Identify Start of Audio

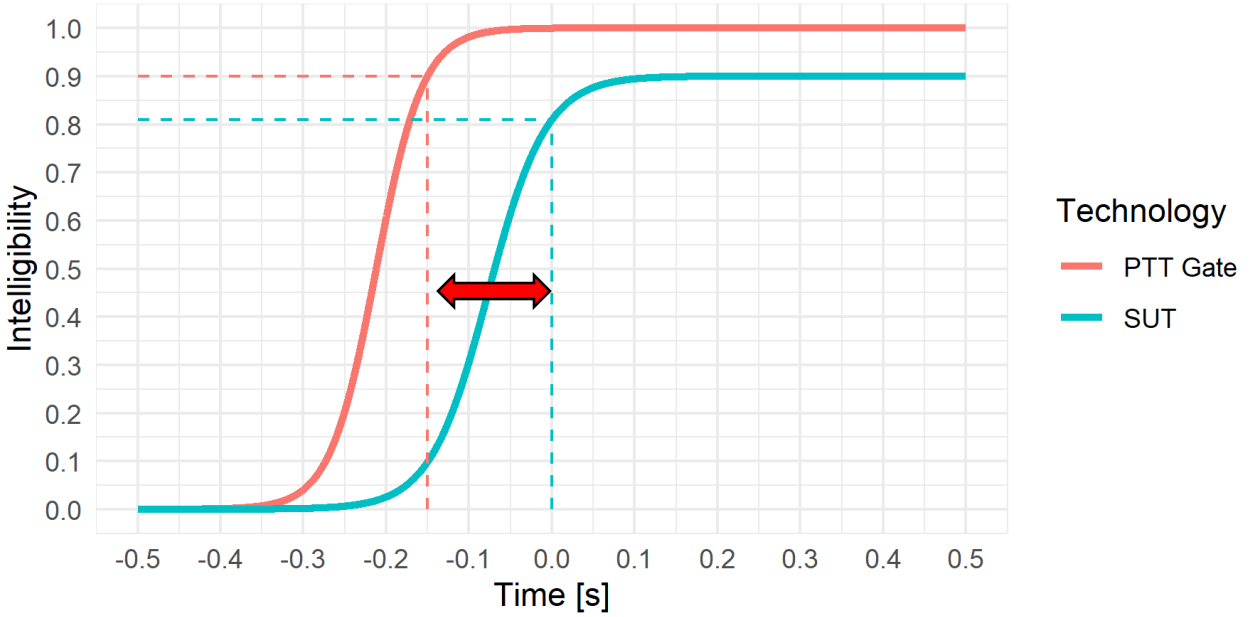


# Start of Word Correction Factor

- **Choose a single word**
- **Fix $\alpha$**
  - $\alpha = 0.9$ for now, 90% intelligibility point
- **Calculate access reference point**
  - Using analog switch (PTT Gate) calculate $\tau_{pg}(\alpha)$
  - How much of the beginning of the keyword clip can be cut
  - PTT Gate access delays are always negative by definition
  - Expendable audio
  - Critical point within a word
- **For a given system-under-test (SUT) calculate $\tau_{sut}(\alpha)$**
  - Time between PTT and starting the keyword clip to achieve 90% intelligibility
- **Calculate true access delay as $\tau(\alpha) = \tau_{sut}(\alpha) - \tau_{pg}(\alpha)$**
  - Time between PTT and keyword clip for 90% intelligibility
  - Corrects for time in keyword clip irrelevant to 90% intelligibility

# Start of Word Correction Factor

- **Fix $\alpha = 0.9$**
- **Think about the word "hook"**
  - Long "h" sound, takes up 200 ms
  - Need 100 ms to understand "hook" from "took", "look", etc…
  - Bad start selection, 50 ms of silence at beginning
- $\tau_{pg}(\alpha) = -150$ ms
- $\tau_{sut}(\alpha) = 0$ ms
  - System required 0 ms between PTT being pressed to send "hook" with 90% of asymptotic intelligibility
  - Currently: no access delay
  - Know that the first 150 ms of the audio clip "worthless"
- $\tau(\alpha) = \tau_{sut}(\alpha) - \tau_{pg}(\alpha) = 0 \ - -150 = 150$ ms
  - Captures the "truth" better
  - System needs 150 ms of setup time to send "hook" through intelligibly



# Future Work

- **Baseline intelligibility**
  - Redo some sessions
- **Access MRT experiment**
- **New speech intelligibility estimation algorithm**
  - Focused on partially muted word impairment

# New MRTs: Design Plan

- **MRT lists**
  - 8 lists
  - 6 words per list
- **Vocoders**
  - Analog
  - P25
  - AMR
- **Cuts**
  - 40 ms intervals

# Creating New Intelligibility Estimation Algorithm

- **Compute accuracy of ABC-MRT16 access delay system**
- **Develop specialized intelligibility estimator**
  - Focused on partially muted words
- **Develop algorithm to differentiate intelligibility on a more fine tuned scale than whole words**
  - One-to-one measurement
  - Requires fine time resolution

# Related Talks

- **PSCR 2020: The Digital Experience**
  - MCV QoE Measurement Methods Presentation
- **2019 Annual Public Safety Broadband Stakeholder Meeting**
  - Mission Critical Voice Quality of Experience Speech-Based Access Measurement

# Related Publications

- Mission Critical Voice Quality of Experience Access Time Measurement Methods
- Mission Critical Voice QoE Mouth-to-Ear Latency Measurement Methods

# THANK YOU

#PSCR2020